January 2006

Evaluating the Biological Condition of Dorset-Area Streams with Rapid Bioassessments

Protecting our environment.

Ontario

2007-05473/

Evaluating the Biological Condition of Dorset-Area Streams with Rapid Bioassessments

Ministry of the Environment

Evaluating the Biological Condition of Dorset-Area Streams with Rapid Bioassessments

Report prepared by:
Keith M. Somers[1], R.A. Reid[1], R.J. Hall[1,2], S. Dosser[3], and B. Clark[1]

[1] Dorset Environmental Science Centre
Environmental Monitoring and Reporting Branch
Ontario Ministry of the Environment
PO Box 39
Dorset, ON P0A 1E0

[2] Present address: Standards Development Branch
Ontario Ministry of the Environment
Rexdale, ON M9P 3V6

[3] Muskoka Lakes Fisheries Assessment Unit
Ontario Ministry of Natural Resources
Bracebridge, ON P1L 1W9

Cette publication technique
n'est disponible qu'en anglais

Printed on recycled paper

Copyright ©: Queen's Printer for Ontario, 2006
This publication may be reproduced for non-commercial purposes with appropriate attribution.

ISBN 1-4249-0112-X

PIBS 5428e

EXECUTIVE SUMMARY

The reference-condition approach is gaining popularity as a method for assessing environmental impacts when traditional upstream-downstream, or before-after comparisons are not feasible. In this study, we used rapid bioassessment data from 10 relatively unimpacted, south-central Ontario streams to estimate biological conditions (i.e., benthic invertebrate communities) in reference streams. With these data, we used the concept of the normal range of variation to generate trial regional reference values for 20 summary biological indices. Data from 6 Dorset-area streams with a known history of water-quality impairment were used to evaluate the regional reference values. Simple comparisons indicated that at least 1 of the 20 biological indices was unusual with respect to the regional reference values for 5 of the 6 Dorset-area streams. Formal statistical evaluations using multivariate analysis of variance (MANOVA) revealed that the benthic communities of all 6 test streams were extremely unusual with respect to the reference streams. Discriminant functions analysis (DFA) was used to identify indices that were most important for distinguishing each test stream from the reference streams.

Our results demonstrate that rapid bioassessment data can be used to characterize benthic invertebrate communities and that the reference-condition approach can be used to identify potentially impacted sites. Because the reference-condition approach does not provide a single, numerical summary of the biological condition of a test site, we advocate the use of MANOVA and DFA to summarize a variety of biological indices and statistically evaluate the relative condition of each test site. These multivariate methods exhibited greater sensitivity relative to simple comparative approaches, and provided a single statistical evaluation of the biological condition of each test stream. Although rapid bioassessment data are often assumed to be relatively insensitive for detecting unusual biological condition, our use of a variety of summary indices and statistically rigorous multivariate methods suggests that this assumption is unwarranted.

TABLE OF CONTENTS

Executive Summary i

Table of Contents ii

List of Tables iii

List of Figures iii

Introduction 1

Methods 2
- Study streams 2
- Sampling protocols 4
- Characterizing the normal range 6
- Statistically judging impairment 7

Results 8
- Water chemistry 8
- Rapid bioassessments 9
- Evaluating the test streams 12
- Multivariate assessment 13
- Judging impairment 15

Discussion 17
- Are regional reference values useful? 20
- Evaluating multiple indices simultaneously 22

Conclusions and Recommendations 24

Acknowledgements 24

Literature Cited 24

LIST OF TABLES

Table 1 Range, median and quartiles of the water-chemistry parameters for 10 reference streams and observed values for the 6 test streams arranged in ascending order of pH 9

Table 2 Regional-reference values based on the distributions of 20 biological indices from 10 south-central Ontario streams. 11

Table 3 Observed values, standardized differences (δ) and generalized distances for the biological indices from the 6 test streams. 16

Table 4 Univariate *t* values and probabilities, plus multivariate partial T^2 values for the 11 two-tailed indices associated with each test stream discriminant analysis 18

LIST OF FIGURES

Figure 1 Locations of the 10 reference streams surrounding the Muskoka lakes in south-central Ontario 3

Figure 2 Rapid bioassessment bench sheet designed for common taxa found in south-central Ontario lakes and streams 5

Figure 3 Scatterplot of: (A) the scores for the 30 reference samples and 6 test streams; and (B), vectors for the major taxa from a correspondence analysis (CA) ordination of the log-transformed relative abundance data. The proportion of the total variance summarized by each axis is indicated by the value in brackets 12

Figure 4 Scatterplot of: (A) the scores for the 30 reference samples and 6 test streams; and (B), vectors for the major taxa; from a correspondence analysis (CA) ordination of the presence - absence data. The proportion of the total variance summarized by each axis is indicated by the value in brackets 13

INTRODUCTION

Despite the large number of environmental impact studies, data on relatively unimpacted reference areas are rare (Hughes et al. 1986, Aagaard et al. 1997, Reynoldson et al. 1997). With no description of reference conditions, objective benchmarks are wanting (Billheimer et al. 1997), the assessment of potentially impacted sites is challenging (Underwood 1992, 1994, Hughes 1995, Glasby 1997), and targets for remediation or rehabilitation are ambiguous (Hartig et al. 1997, Yan et al. 1996). At present, there is no consensus on the best way to characterize unimpacted areas (Barbour et al. 1995, Reynoldson et al. 1997, Roux et al. 1999).

Given the general lack of biological criteria (Barbour et al. 1995, 1996) and the growing realization that biological criteria should be based on observed values at unimpacted, or minimally impacted, reference areas (Hughes 1995, Hartig et al. 1997), a comparative assessment technique known as the reference-condition approach is gaining popularity (Hughes et al. 1986, Reynoldson et al. 1997, Bailey et al. 1998). The reference-condition approach utilizes natural variability within and among reference sites in order to define biological expectations or targets (e.g., Underwood 1992, 1994, Osenberg et al. 1994, Glasby 1997). Biological criteria based on this approach are often called regional reference values (Hughes et al. 1986), and biological condition is evaluated by simple comparisons between the observed state at a test site and the regional reference values (Reynoldson et al. 1997, Bailey et al. 1998, David et al. 1998).

Kilgour et al. (1998) recently described a statistical approach for assessing a potentially impacted site using criteria based on the normal range of variation of reference sites. They defined the normal range as the range enclosing 95% of the data from reference-area sites (e.g., Leffler 1978, Kersting 1991). Test sites falling within this range are judged to be biologically healthy (i.e., unimpacted), whereas sites lying outside of the normal range are atypical, and potentially impacted. Although most assessments using reference sites do not involve formal statistical tests (e.g., Barbour et al. 1996, Reynoldson et al. 1997), Kilgour et al. (1998) provide formulae for statistically evaluating if a test site falls outside of the normal range. Traditional statistical methods test the null hypothesis that the observed biological difference (or distance) between the test site and the reference-site mean ($\bar{x}$) is not significantly different from zero (e.g., McBride et al. 1993). By contrast, Kilgour et al. (1998) test the hypothesis that the observed difference exceeds some critical non-zero difference (or effect size) that is based on the normal range. Kilgour et al. (1998) also propose that the normal range by definition provides a criterion for tentatively defining an ecologically significant impact. That is, a test site that falls outside of the normal range suggests an ecologically significant difference, not just a statistically significant difference. However, their proposal that the normal range defines the threshold for an ecologically significant impact is untested.

Many softwater streams on the Precambrian Shield of south-central Ontario, Canada have been impacted by acid deposition (LaZerte and Dillon 1984,

Molot et al. 1989, Dillon and LaZerte 1992) and may be threatened by global warming (Schindler et al. 1996, Dillon et al. 1997, Schindler 1998). Unfortunately, our knowledge of the macroinvertebrate communities of softwater streams is limited (Mackay and Kersey 1985, Hall and Ide 1987, Chmielewski and Hall 1993). The goals of this study were to (1) summarize rapid bioassessment data on benthic macroinvertebrate communities from 10 south-central Ontario streams and characterize the normal range of variation for relatively unimpacted streams in this area; (2) use the resultant normal ranges to assess the biological condition of 6 additional streams with known histories of water-quality impairment associated with acid deposition (Mackay and Kersey 1985, Dillon et al. 1991); and (3) evaluate the comparative assessment of the test streams in (2) using the statistical framework of Kilgour et al. (1998). Also, because both assessments (i.e., 2 and 3) examine summary indices one at a time, and there is no guidance as to how many individual tests must be significant before a site is declared impaired, we present a multivariate statistical approach to judge impairment based on a number of biological indices. Assuming that some of the test streams are identified as being unusual relative to the reference streams, we propose that our tabulated normal ranges can serve as trial regional reference values (or benchmarks) to assess the impact of future disturbances operating at both local and regional geographic scales in south-central Ontario (e.g., Dillon et al. 1997, Clair et al. 1998, Schindler 1998).

METHODS

Rapid bioassessment methods are simple inexpensive ways to quickly assess the biological condition of streams (Lenat and Barbour 1994, Chessman 1995, Resh et al. 1995). In rapid bioassessments, the costs of traditional biological assessments are reduced by using kick-and-sweep sampling, removing only a small fraction of the total number of individuals (e.g., 100 animals), and identifying the individuals to a coarse taxonomic level (e.g., order; see Plafkin et al. 1989, Culp et al. 1997, David et al. 1998). Despite this reduction in scientific rigor, rapid bioassessment methods are capable of screening sites into relatively coarse categories of biological condition (Barbour et al. 1992, Hannaford and Resh 1995, Resh et al. 1995, Growns et al. 1997). However, the choice of summary biological indices and the most appropriate ways to evaluate test sites are still debated (Barbour et al. 1995, Gerritsen 1995, Norris 1995, Yoder and Rankin 1995). Perhaps because of marked regional differences in benthic macroinvertebrate communities (e.g., Corkum 1989), there is a growing consensus within these debates that assessment criteria should be based on a collection of relatively unimpacted reference sites from the general vicinity of the test site (Hughes 1995, Glasby 1997).

Study streams -Ten streams that flow into lakes Joseph, Rosseau and Muskoka were selected to characterize the normal range of variation of relatively unimpacted streams in south-central Ontario, Canada (Fig. 1). Five of these streams are 2nd or 3rd order streams, draining catchments containing multiple small lakes. The other 5 are smaller streams draining lakes in relatively small catchments. These 10

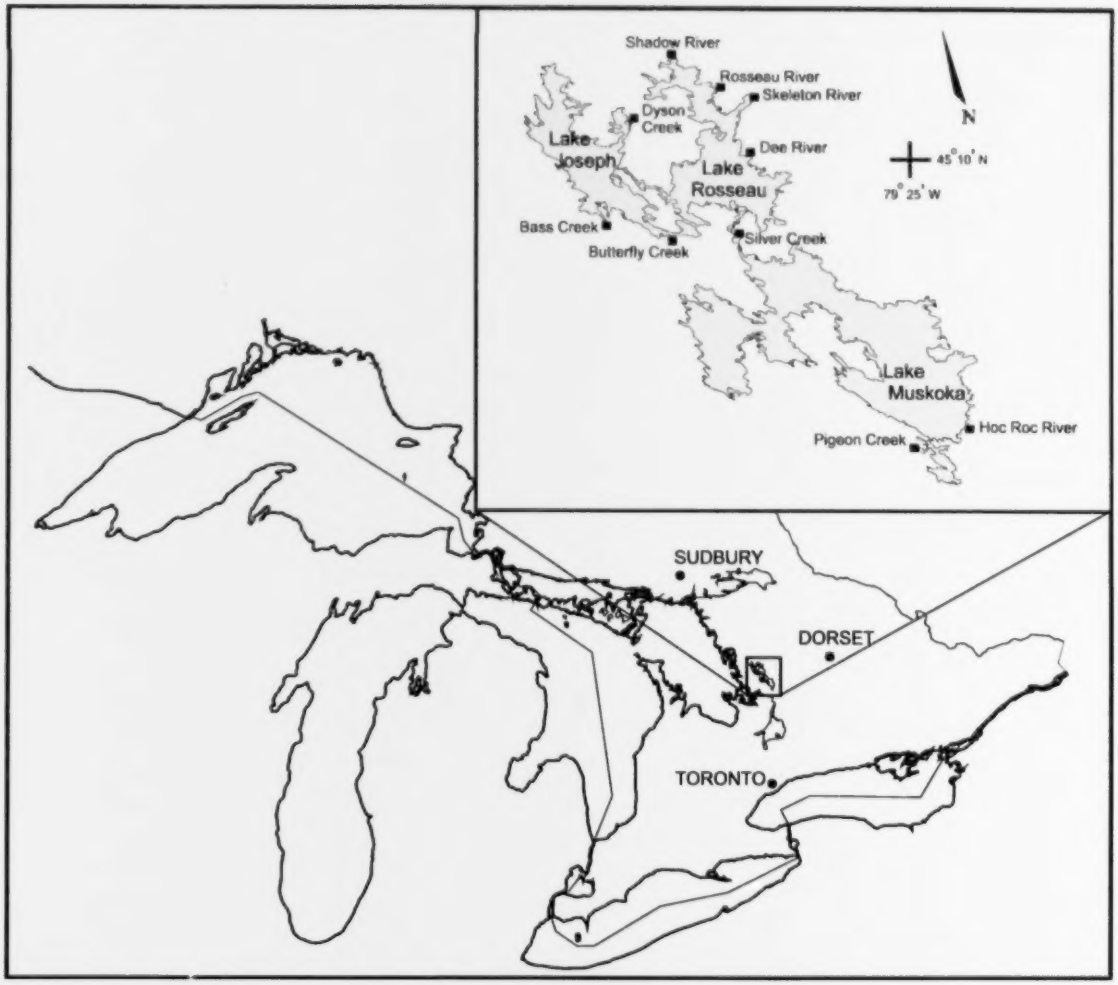

Figure 1: Locations of the 10 reference streams surrounding the Muskoka Lakes in south-central Ontario

reference streams are within 40 km of each other.

Six unnamed streams within 25 km of the Dorset Environmental Science Centre were selected as test streams to evaluate the benchmarks that were based on the 10 reference streams. In 1981, the invertebrate communities of 5 of these streams were thought to be impacted by water-quality impairment due to acid deposition (Mackay and Kersey 1985). All 6 test streams are small first or second order streams that are approximately 100m higher in elevation than the reference streams. The test streams are within 35km of each other and are located <70 km to the east of the reference streams. All 16 streams are typical of the Precambrian Shield and are unique

because of shallow soils comprised of glacial till, numerous wetlands and lakes, and bedrock ridges (Jeffries and Snyder 1983, Dillon et al. 1991). However, there are marked differences in the thickness of the soils in the catchments of the test and reference streams. The Dorset-area streams lie in catchments with relatively thin soils (i.e., generally <3 m, Jeffries and Snyder 1983, Dillon and Molot 1997), whereas the reference streams are in catchments with thicker soils. Presumably this difference in soil thickness is one of the reasons why the test streams are influenced by acid precipitation (and precipitation chemistry) to a greater degree than the reference streams (Dillon and Molot 1997).

More than two decades of water-chemistry data for Dorset-area streams underscore the long-term impacts of acid deposition on stream pH and associated chemical parameters (LaZerte and Dillon 1984, Mackay and Kersey 1985). With the legislated reductions in atmospheric emissions, the water chemistry of many streams has gradually improved over the last decade (Dillon and LaZerte 1992), although several unusually dry years have recently resulted in slow recovery (Dillon et al. 1997).

Although there are some surveys of benthic communities in some of the Dorset-area streams (Mackay and Kersey 1985, Hall and Ide 1987, Chmielewski and Hall 1993), no simple model relating biological condition to water quality exists. Consequently, we are unable to predict the macroinvertebrate community of each test stream given the water chemistry. However, the benthic communities in the test streams appear to be recovering from historical impairment, and given the acidification history of these streams (LaZerte and Dillon 1984, Dillon and LaZerte 1992), we expect the most unusual benthic communities to be found in the streams with the lowest pH.

Sampling protocols - Each stream was sampled in mid-May 1995 after high spring flows had receded. Samples of water were collected using standard protocols (Girard and Reid 1990, Clark 1996), and a series of water-chemistry parameters were measured using standard analytical methods (Anonymous 1983). Gran alkalinity, pH, chloride concentration, conductivity, dissolved inorganic carbon (DIC), dissolved organic carbon (DOC), total Kjeldahl nitrogen (TKN), total phosphorus, and sulphate concentrations were tabulated to compare the 10 reference streams and the 6 test streams.

Benthic macroinvertebrates were sampled using a rapid-bioassessment protocol (David et al. 1998). Three samples were collected from shallow riffle areas in each stream. A 1-m^2 area was randomly selected within a given riffle and the substrate and associated invertebrates were collected with a kick-and-sweep technique using a 250-µm-mesh D net. Each of the three 1-m^2 areas was sampled for one minute. The debris and associated invertebrates were rinsed in a 1000-µm-mesh sieve to remove the fine sand and then preserved in 10% buffered formalin. The preserved samples were taken to a laboratory where randomly selected subsamples were sorted until at least 100 macroinvertebrates were enumerated. On occasion, the entire sample was sorted without obtaining a total of 100 animals. The resultant macro-invertebrates were identified to a coarse taxonomic level, and enumerated on

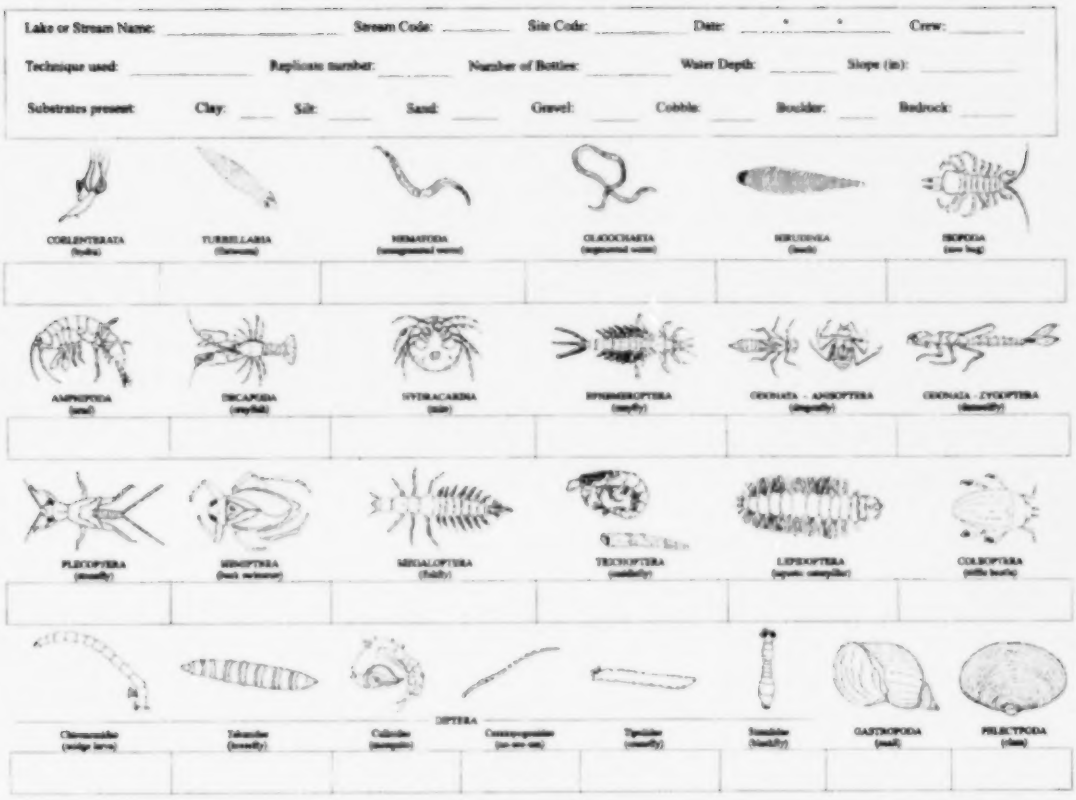

Figure 2: Rapid bioassessment bench sheet designed with common taxa found in south-central Ontario lakes and streams.

bench sheets (Fig. 2). The final data matrix consisted of counts of the number of individuals of the various taxa in the 3 replicate samples from each of the 10 reference streams and $\bar{x}$ values for each of the 6 test streams. We did not average replicate samples from the reference streams in order to include both within- and among-stream variability in our estimate of reference-stream variation. By contrast, we used $\bar{x}$ values for each test stream to obtain a single, best estimate of biological condition and to simplify the subsequent presentation of our results. We recognise that more appropriate experimental designs would be preferred if this were a formal assessment, instead of a demonstration.

The counts of the different taxonomic groups were summarized using a wide range of richness, compositional, and multivariate indices. Simple indices included: (i)Four richness metrics: 1) taxonomic richness (i.e., number of taxonomic groups, see Fig. 2); 2) dipteran richness, the number of dipteran families; 3) non-insect richness, the number of taxonomic groups that were not insects; and 4) non-dipteran insect richness, the number of groups represented by non-dipteran insects; and (ii)10 compositional metrics based on the number of individuals in each sample: 1) %

Ephemeroptera, Plecoptera, and Trichoptera (EPT); 2) % non-dipteran insects; 3) EPT/chironomid ratio; 4) % oligochaetes; 5) % dominants (the proportion of the total represented by the group with the maximum number of individuals); 6) % chironomids; 7) % crustaceans; 8) % dipterans; 9) % gastropods; and 10) % pelecypods. Low values of the 4 richness indices and the first 3 compositional metrics, high values of (4) and (5), and both low and high values of the remaining 5 compositional metrics were assumed to indicate potentially impaired benthic communities (e.g., Plafkin et al. 1989, Barbour et al. 1995, Reynoldson et al. 1997).

The stream-by-sample-by-taxon matrix of counts was also used to calculate six multivariate metrics. Counts were logarithmically transformed to downweight the most abundant taxa (i.e., log[X+1]) and then summarized using a correspondence analysis (CA) ordination (Rohlf 1993). The first 3 axes from the CA were retained as multivariate, composition-type metrics. The counts were also recoded to simple presence (1) or absence (0), and the first 3 axes from a second CA were used as 3 multivariate richness metrics.

Because the $\bar{x}$ values for the test streams were based on counts of a total of 300 animals and the individual (or replicate) samples for the reference streams were based on counts of only 100 animals, all simple richness metrics were calculated for the individual replicates for each test stream. The resultant 3 values were averaged to obtain a single estimate for each test site. These calculations were necessary because richness metrics are influenced by the total number of animals counted (Fisher et al. 1943, Courtemanch 1996), and a comparison of richness metrics based on 100- and 300-count samples would undoubtedly lead to the detection of differences that were sample-size artifacts. A similar issue could arise in the CA ordination of the presence-absence data. As a result for the ordination, presence in a test stream was only recorded if a taxon was found in 2-or-more of the 3 samples.

Rare taxa were eliminated from the data matrix prior to the ordinations because rare taxa can have undue influence in CA (ter Braak and Prentice 1988). Rare taxa were arbitrarily defined as those taxonomic groups found in <3 sites or represented by <10 individuals in the full data matrix. All 6 multivariate indices were assumed to be 2-tailed, with extreme values at either end of an axis indicating potentially impaired benthic communities (e.g., Warwick and Clarke 1991, Reynoldson et al. 1997).

Characterizing the normal range - Kilgour et al. (1998) described a simple way to characterize the normal range of variation among samples from reference areas. For any summary index, $\bar{x} \pm 1$ Standard Deviation (SD) encloses 68% of the samples, assuming that the index is normally distributed. Similarly, $\bar{x} \pm 2$ SDs includes 95% of the values. Thus, Kilgour et al. defined the normal range as the $\bar{x} \pm 2$ SDs, or the 95% confidence region for any index based on reference-area data. Note their use of the SD not the Standard Error (SE) because the normal range is intended to enclose 95% of the individual data points, not the 95% confidence region for the $\bar{x}$.

David et al. (1998) used a rank-order variant of the normal range to define regional reference values. Given a set of

reference-area samples, the frequency distribution for each biological index was summarized in percentiles. Depending on the type of index (i.e., 1-tailed metrics are indices where either high or low values indicate impairment; 2-tailed metrics are indices where both high and low values indicate impairment), critical values were defined as the 95th percentile for 1-tailed metrics (or 5th percentile), or the 2.5 and 97.5 percentiles for 2-tailed indices. To provide additional guidance, David et al. distinguished samples lying outside of the normal range, but enclosed by the 99.9th percentile as atypical (i.e., between 2 and 3 SDs from the $\bar{x}$). Samples falling beyond the 99.9th percentile were considered to be extremely unusual. Using this construct, 5 of 100 random samples from reference areas would be incorrectly classified as lying outside of the normal range. Similarly, 1 in 1000 random reference samples would be identified as an extreme sample. We used this rank-order approach to define the normal range of variation for the 10 reference streams and the critical percentiles were tabulated as tentative regional reference values for softwater streams in south-central Ontario.

Regional reference values were used to evaluate the biological condition of the six test streams. Because we were interested in the overall biological condition of each test stream, a single estimate for each index was compared to the regional reference values in order to classify each test stream as normal, atypical or extreme. We counted the number of extreme and atypical values as a simple estimate of biological condition. To further quantify the condition of each test stream, we calculated the difference between the test-stream value and the reference-stream $\bar{x}$ for each metric and standardized this difference by dividing by the reference-stream SD (i.e., $[x_{test} - \bar{x}_{ref}] / SD_{ref}$ as in Diamond et al. [1996] and Kilgour et al. [1998]). We tabulated these standardized differences (i.e., δs with units in SDs) to facilitate comparisons among test streams and among indices. To be consistent with the regional reference values, δs >3 SDs were considered to be extreme, whereas δs between 2 and 3 SDs were judged to be atypical.

Statistically judging impairment - Because each assessment involved 20 indices, we wanted a summary that provided a single estimate of the biological condition of each test stream. Additive multimetric summaries (e.g., Barbour et al. 1995, Gerritsen 1995) and multivariate methods (e.g., Norris 1995, Reynoldson et al. 1997) are often used for this task. We chose a multivariate approach and calculated generalized (or Mahalanobis) distances between each test stream and the $\bar{x}$ (or centroid) of the reference streams using our various biological indices (e.g., Kilgour et al. 1998). One-tailed metrics were not included in the calculations because the generalized distance treats extreme values at both ends of an index as equally deviant. The generalized distance is a summary metric that accounts for correlations between indices, such that redundant information associated with correlated indices is down-weighted in the calculations (Sneath and Sokal 1973, Manly 1986). As a result, the generalized distance is not influenced if we select a number of correlated indices.

The generalized distance between the reference stream centroid and a given test stream is an extension of a multivariate t-test (Hotelling's T^2) that is

readily evaluated as a multivariate analysis of variance (MANOVA) with an F value and associated probability (Manly 1986). We used the generalized distance to test 2 hypotheses for each test stream: 1) that the observed distance was not significantly different from 0 (i.e., the traditional test of statistical significance); and 2), that the test stream was not significantly outside of the normal range (i.e., the non-central test described by Kilgour et al. 1998).

To complement the MANOVA, we used discriminant functions analysis (DFA) to identify those indices that best separated a given test stream from the reference-stream centroid. Individual DF axes were calculated for each test stream (i.e., 6 separate DFAs) and the most important indices were identified by comparing univariate t values and multivariate partial T^2 values associated with each DF axis (Rencher and Scott 1990, Rencher 1992). The univariate t values summarize the results of one-sample t tests comparing the test stream to the $\bar{x}$ of the reference streams, 1 index at a time. These t values are analogous to correlations between a given index and the DF axis (Rencher 1992), such that the largest t values reveal those indices that are most strongly correlated with the DF axis. By contrast, the partial T^2 values reveal unique information contributed by each index given all of the other indices (Rencher and Scott 1990). Thus, the partial T^2 indicates how each index uniquely contributes to the unusual nature of the test stream after controlling for correlations with the other indices.

We also used the distribution of generalized distances between the reference-area centroid and each of the 30 reference samples to estimate the 95th and 99.9th percentiles as multivariate benchmarks. These percentiles are multivariate analogues of the percentile-based regional reference values. Other assessment methods, such as the percent model affinity approach (Novak and Bode 1992) could be modified in a similar way to use site-specific reference samples (e.g., Barton 1996).

RESULTS

Water chemistry - We found marked differences in the water chemistry of the reference and test streams that are consistent with differences in acid sensitivity (Table 1). The pH of the reference streams ranged from 5.9 to 7.0, with a median value of 6.7. By contrast, the pH of all 6 test streams was <5.9, and Gran alkalinity was always less than the median for the reference streams. Sulphate concentrations in the 3 test streams with the lowest pH were below the minimum value for the reference streams. By contrast, sulphate concentrations for the other test streams exceeded the reference-stream 75th percentile. The 3 test streams with the lowest pH also had the highest concentrations of DOC reflecting the fact that these streams were in catchments with sizeable wetlands (i.e., >10% of the area, Dillon and Molot 1997).

Concentrations of nutrients and DIC in the test streams generally fell within the range of values for the 10 reference streams (Table 1). However, chlorine (Cl) concentrations and conductivity for the test streams near Harp Lake, HP3 and HP6, were noticeably higher than the other test streams, which generally had

Table 1. Range, median and quartiles of the water-chemistry parameters for 10 reference streams and observed values for the 6 test streams.

Water-chemistry	Reference streams					Test streams					
parameter	Minimum	25th Percentile	Median	75th Percentile	Maximum	DE11	PC1	DE6	CB1	HP3	HP6
Gran alkalinity (mg/L $CaCO_3$)	1.91	3.51	4.84	8.13	9.63	-0.79	-0.07	1.15	3.30	3.83	4.49
pH	5.92	6.42	6.68	6.78	6.96	4.36	4.47	4.75	5.55	5.80	5.83
Chloride (mg/L)	0.51	1.83	3.24	7.71	19.80	0.16	0.07	1.60	0.21	6.48	13.40
Conductivity (uS)	24.5	27.4	39.7	60.0	105.0	23.15	20.90	18.65	22.70	46.10	74.10
DIC (mg/L)	0.79	1.15	1.46	2.21	2.57	2.47	0.98	2.70	2.17	1.36	2.28
DOC (mg/L)	1.8	3.9	5.0	7.1	7.8	16.8	11.9	13.8	6.5	8.3	6.3
TKN (ug/L)	160	285	340	410	460	520	380	520	280	360	320
Total phosphorus (ug/L)	6.4	8.2	13.4	18.1	38.6	24.5	7.2	35.4	7.2	13.0	12.2
Sulphate (mg/L SO_4)	3.90	4.04	4.50	5.06	7.05	2.73	3.75	0.90	5.75	5.40	6.70

values below the reference-stream minimum. The higher Cl and conductivity values probably reflect the use of road salt within these 2 catchments during the winter.

Rapid bioassessments - All reference and test streams were sampled from May 15 - 18, 1995. Two samples from one of the reference creeks were lost and replaced by samples collected on September 21, 1995. Most indices for these 2 replacement samples fell within the normal range. However, 1 replacement sample, containing 40% oligochaetes, fell between the 95 and 97.5 percentiles for the reference streams (Table 2; the most extreme reference sample was 79.3%). The 2nd replacement sample contained 81% chironomids. This large number of chironomids was unusual (although the next highest value was 78.8%), resulting in extreme values for % chironomids, % dipterans, and % dominants. Despite these unusually high values, none of the multivariate metrics indicated that the two replacement samples were atypical. As a result, we included the replacement samples when calculating the normal range of variation for the reference-area streams.

Although our rapid bioassessment protocol advocates counting 100 animals from each sample (David et al. 1998, Somers et al. 1998), 5 of 30 samples sorted completely without obtaining 100 animals (i.e., producing counts of 36, 40, 83, 65, and 77). Total counts for the remaining reference samples ranged from 92 to 122, whereas counts for the 18 samples from the six test streams ranged from 99 to 107 animals.

Separate CA ordinations of the log-transformed abundance data and the presence-absence (P/A) data produced similar results. The abundance CA summarized 61.3% of the variation in the taxa-by-sample matrix in the first 3 axes. A plot of the 1st and 2nd axes revealed a distinct elliptical cluster of 26 of 30 reference samples (Fig. 3A). The major axis of this ellipse was based on high proportions of Isopoda and/or Amphipoda at the positive end of CA axis 2, and high

Table 2. Regional-reference values based on the distributions of 20 biological indices from 10 south-central Ontario streams. Bold values indicate critical values outside of the 95th percentile and bold with italics indicate critical values outside of the 99.9th percentile.

Index	Percentiles for the reference-stream frequency distributions														
	0.05	0.1	1	2.5	5	10	25	50	75	90	95	97.5	99	99.9	99.95
Taxonomic richness	***5.0***	***5.0***	**5.0**	**5.0**	**5.0**	5.0	6.0	7.0	8.0	9.1	10.6	11.0	11.0	11.0	11.0
Dipteran richness	***1.0***	***1.0***	**1.0**	**1.0**	**1.0**	1.0	2.0	2.0	2.0	2.0	2.0	2.3	2.7	3.0	3.0
Non-insect richness	***1.0***	***1.0***	**1.3**	**1.7**	**2.0**	2.0	3.0	4.0	5.0	5.0	5.6	6.0	6.0	6.0	6.0
Non-dipteran insect richness	***0.0***	***0.0***	**0.0**	**0.0**	**0.0**	0.0	1.0	1.0	2.0	3.0	3.6	4.3	4.7	5.0	5.0
% EPT	***0.00***	***0.00***	**0.00**	**0.00**	**0.00**	0.00	0.20	2.42	17.08	31.53	37.27	41.42	45.19	47.44	47.57
% non-dipteran insects	***0.00***	***0.00***	**0.00**	**0.00**	**0.00**	0.00	0.89	2.91	19.01	35.18	37.61	41.85	46.28	48.94	49.06
EPT/chironomids	***0.00***	***0.00***	**0.00**	**0.00**	**0.00**	0.00	0.00	0.06	0.49	1.17	1.63	4.21	7.75	9.88	10.00
% oligochaetes	0.00	0.00	0.00	0.00	0.48	1.07	5.41	10.00	21.84	28.62	**38.05**	**50.95**	**67.93**	***78.11***	***78.68***
% dominants	26.28	26.30	26.62	27.16	28.63	31.67	37.81	46.78	55.34	73.41	**79.04**	**79.71**	**80.46**	***80.90***	***80.93***
% chironomids	***4.64***	**4.67**	**5.15**	**5.95**	10.34	15.21	24.49	36.58	49.77	62.28	76.10	**79.38**	**80.32**	**80.89**	***80.92***
% crustaceans	***0.00***	**0.00**	**0.00**	**0.00**	0.00	0.00	0.24	6.63	20.33	50.22	61.80	**67.59**	**71.90**	**74.49**	***74.63***
% dipterans	***6.58***	**6.71**	**9.04**	**12.93**	16.29	17.87	30.43	44.92	53.68	72.28	78.88	**84.88**	**86.52**	**87.51**	***87.56***
% gastropods	***0.00***	**0.00**	**0.00**	**0.00**	0.00	0.00	0.00	1.43	6.40	26.19	29.88	**37.24**	**44.57**	**48.97**	***49.22***
% pelecypods	***0.00***	**0.00**	**0.00**	**0.00**	0.00	0.00	0.97	3.24	7.59	14.46	15.92	**20.96**	**27.50**	**31.42**	***31.64***
Abundance CA axis 1	***-0.796***	**-0.795**	**-0.778**	**-0.750**	-0.697	-0.647	-0.550	-0.316	-0.084	0.474	0.596	**0.653**	**0.655**	**0.656**	***0.656***
Abundance CA axis 2	***-0.821***	**-0.818**	**-0.766**	**-0.679**	-0.611	-0.585	-0.307	-0.179	0.290	0.606	0.655	**0.709**	**0.783**	**0.828**	***0.831***
Abundance CA axis 3	***-1.167***	**-1.160**	**-1.031**	**-0.817**	-0.656	-0.339	-0.105	0.055	0.213	0.363	0.545	**0.706**	**0.749**	**0.775**	***0.776***
P/A CA axis 1	***-0.729***	**-0.728**	**-0.710**	**-0.680**	-0.649	-0.635	-0.487	-0.324	-0.073	0.448	0.560	**0.650**	**0.739**	**0.788**	***0.790***
P/A CA axis 2	***-1.003***	**-1.001**	**-0.965**	**-0.906**	-0.757	-0.588	-0.164	0.038	0.087	0.430	0.518	**0.608**	**0.726**	**0.796**	***0.800***
P/A CA axis 3	***-0.562***	**-0.561**	**-0.534**	**-0.490**	-0.460	-0.410	-0.229	-0.052	0.271	0.403	0.723	**0.835**	**0.902**	**0.942**	***0.944***

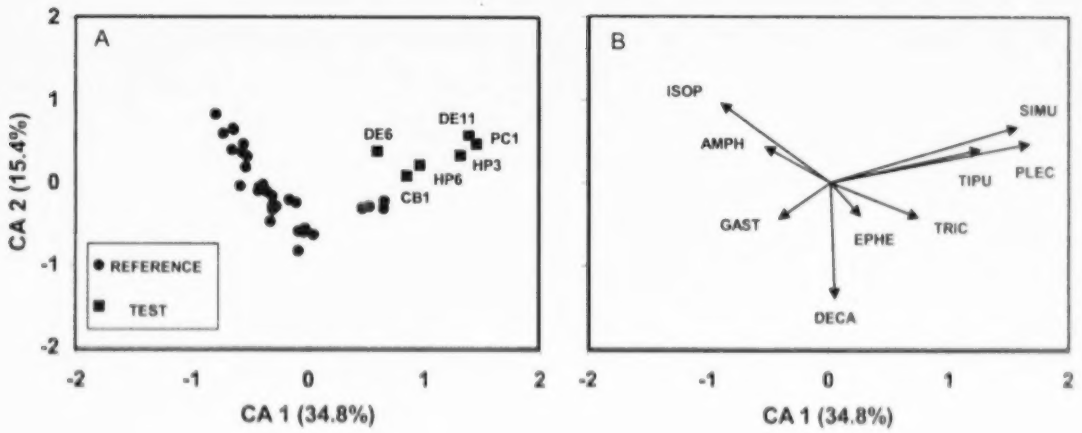

Figure 3: Scatterplot of : (A) the scores for the 30 reference samples and 6 test streams; and (B) vectors for the major taxa from a correspondence analysis (CA) ordination of the log-transformed relative abundance data. The proportion of the total variance summarized by each axis is indicated by the value in brackets.

proportions of Ephemeroptera, Trichoptera, and/or Decapoda at the opposite end (Fig. 3B). All six test streams and four of the reference-area samples fell on a vector at right angles to the major axis of the ellipse. This vector was based on high proportions of Simuliidae, Plecoptera, and/or Tipulidae (Fig. 3B). Three of the four outlying reference samples were from the Skeleton River. These 3 samples had the highest numbers of Plecoptera or Tipulidae of the reference samples. The 4th outlying sample was from Pigeon Creek with a total count of only 65 animals. This sample was the only reference sample that contained Simuliidae (i.e., 8% of the total). In contrast, the test streams averaged 14-36% Simuliidae, and 7-49% Plecoptera, with the exception of one of the test sites, DE6, where Plecoptera were quite rare (i.e., ~1%). Tipulidae were found in 5 of the test streams (not CB1), whereas Gastropoda and Isopoda were generally absent, and Amphipoda were rare (i.e., ~1% in 5 of the 6 test streams).

The P/A CA summarized 62.4% of the variation in the first 3 axes. The 1st and 2nd axes grouped the same 26 reference-area samples together (Fig. 4A), although the resultant cluster of samples exhibited more scatter than in the abundance CA (Fig. 3A). The greatest variation among the 26 reference-area samples was along axis 2. Samples at the positive end of axis 2 were distinguished by the presence of Trombidiformes (i.e., Hydracarina) and/or Isopoda (Fig. 4B), whereas samples with negative scores contained Decapoda and/or Trichoptera. The 6 test streams and 4 unusual reference-area

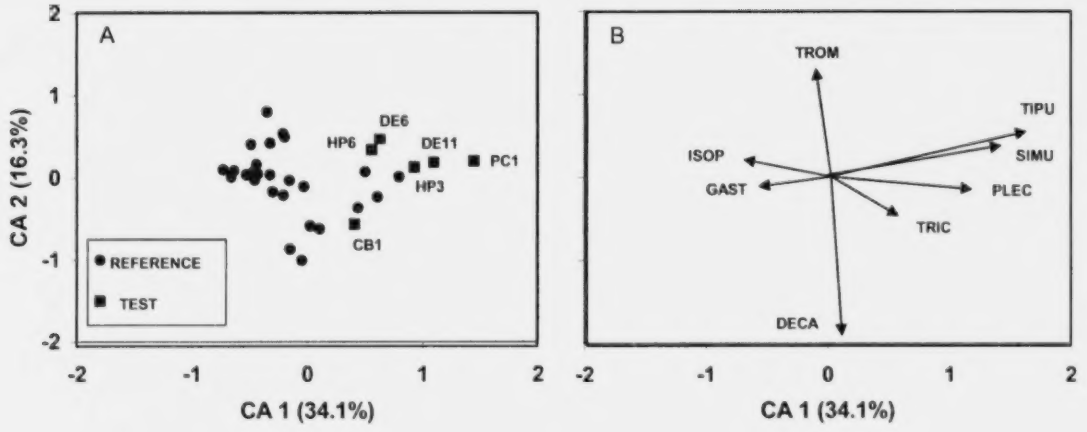

Figure 4: Scatterplot of : (A) the scores for the 30 reference samples and 6 test streams; and (B), vectors for the major taxa; from a correspondence analysis (CA) ordination of the presence-absense data. The proportion of the total variance summarized by each axis is indicated by the value in brackets.

samples were separated from the other 26 samples along P/A CA axis 1 because of the presence of Tipulidae, Simuliidae, and/or Plecoptera, plus the absence of Gastropoda and/or Isopoda. The overall similarity between the 2 ordinations reflects the fact that the major differences between the reference and test streams were a function of the presence or absence of specific taxonomic groups. However, the reduced separation between the test and reference streams in the P/A CA suggests that relative abundance also contributed information that distinguished the communities in the test and reference streams.

Frequency distributions for each metric based on the 30 reference samples were used to estimate regional reference values (Table 2). For 9 of these indices critical percentiles could not be distinguished from percentiles lying within the normal range (i.e. taxonomic, dipteran, and non-dipteran insect richness; %EPT, %non-dipteran insects, EPT/chironomids, % crustaceans, % gastropods, and % pelecycpods). As a result, low values for these 9 indices could not be used to identify potentially impacted sites (a problem we attribute to a truncated distribution that might be resolved with a larger number of reference-area samples). By contrast, 1-tailed metrics where high values indicated a degraded community (i.e., %oligochaetes and % dominants), and the remaining 2-tailed indices (i.e., % chironomids, % dipterans, and the 6 ordination axes) provided unique benchmarks for regional-reference values.

Evaluating the test streams - The percentile-based regional reference values indicated that the test streams varied with respect to their degree of potential impairment, ranging from no impairment to extreme values for a maximum of 3 indices (Table 3). The

biological condition of the test streams could not be evaluated with 3 metrics (% crustaceans, % gastropods, and % pelecypods) because the reference-stream distributions did not provide unique critical values (e.g., the 10th, 5th, and 0.1 percentiles had the same value; Table 2). For these 3 indices, all 6 test streams were assumed to fall within the normal range because the δs were all <1 SD from the reference-area $\bar{x}$ (Table 3).

As expected, the stream with the lowest pH was most unusual biologically with extreme values for 3 metrics (i.e., non-insect richness; and abundance and P/A CA axis 1 at DE11, Table 3). However, the biological condition of the remaining test streams was not consistent with the pH gradient. For example, PC1 with the 2nd lowest pH produced 2 extreme values and 1 atypical value, whereas DE6 exhibited no biological impairment despite having the 3rd lowest pH. By contrast, CB1 was impaired (1 extreme value), as was HP3 with 2 extreme values and 1 atypical value. The stream with the highest pH, HP6, exhibited 1 extreme value.

Assessments using standardized differences (δs) identified the same metric values, in the same sites, as being atypical as did the percentile-based regional reference values (Table 3). In addition, the δs identified two atypical values for DE6 and one additional atypical value for HP6. Because these 3 δs were close to the critical value of 2 SDs, the disagreement between the two assessments may be due to non-normality. Shapiro-Wilks tests indicated that the reference-stream distributions for these indices were significantly non-normal ($p<0.01$, e.g., see Figs 3A, 4A). Thus, the different results for the two assessment methods are probably due to a violation of the normality assumption that underlies our use of the δs.

The δs also revealed a number of extreme values in the tail of the distributions for several indices that were opposite to our expectations (Table 3). For example, dipteran richness was greater than expected in five of the test streams, and taxonomic richness was greater than expected in DE6. In both instances greater richness was not predicted (Table 2), suggesting errors in our assumption that some metrics were 1-tailed (i.e., an impact was assumed to cause a change in only 1 direction).

Multivariate assessment - Generalized distances based on the first 3 axes of the abundance CA indicated that all 6 test streams were unusual biologically (Table 3). That is, all 6 test streams were farther away from the reference-sample centroid than the 99.9th percentile. Based on these distances DE6 was the least impacted, followed by CB1, HP6, HP3, DE11, and PC1. With the exception of the reversal of DE11 and PC1, this assessment ranked the test streams in the same order as the results based on the regional reference values for the individual metrics. By contrast, generalized distances using only the P/A

Table 3. Observed values, standardized differences (δ) and generalized distances for the biological indices from the 6 test streams. The streams are ordered from lowest to highest pH. Atypical values are in bold; extreme values are in bold and italics. Multivariate assessments of different subsets of indices are in the lower panel.

Index	DE11		PC1		DE6		CB1		HP3		HP6	
	Value	δ	Value	δ	Value	δ	Value	δ	Value	δ	Value	δ
Taxonomic richness	7	-0.390	6	-0.596	11	2.076	10	1.459	7	0.021	10	1.459
Dipteran richness	3	2.610	3	1.843	4	4.146	4	4.146	3	3.376	3	3.378
Non-insect richness	***0***	**-2.762**	**1**	**-2.241**	3	-0.417	3	-0.938	**1**	-1.980	2	-1.199
Non-dipteran insect richness	3	1.412	3	0.879	4	1.679	3	1.412	3	0.879	4	1.946
% EPT	45.10	2.522	57.24	3.395	14.87	0.349	31.63	1.554	53.40	3.119	22.44	0.893
% non-dipteran insects	46.41	2.470	58.55	3.321	17.72	0.462	33.23	1.549	54.05	3.005	23.72	0.882
EPT/chironomids	1.22	0.326	6.37	3.111	0.83	0.119	2.52	1.028	3.64	1.635	0.54	-0.042
% oligochaetes	1.63	-0.857	10.53	-0.305	16.14	0.044	2.88	-0.780	0.97	-0.898	1.92	-0.839
% dominants	43.14	-0.354	48.36	-0.010	30.06	-1.217	35.78	-0.839	37.22	-0.745	41.67	-0.451
% chironomids	36.93	-0.058	8.88	-1.469	17.72	-1.024	12.46	-1.289	14.56	-1.183	41.67	0.181
% crustaceans	0.00	-0.719	0.00	-0.719	13.61	-0.095	0.64	-0.690	0.00	-0.719	0.64	-0.690
% dipterans	51.96	0.362	30.26	-0.711	50.95	0.312	59.74	0.747	38.83	-0.287	72.12	1.359
% gastropods	0.00	-0.561	0.00	-0.561	0.00	-0.561	0.00	-0.561	0.00	-0.561	0.32	-0.533
% pelecypods	0.00	-0.827	0.00	-0.827	0.00	-0.827	1.60	-0.599	4.53	-0.180	0.64	-0.736
Abundance CA axis 1	***1.388***	***4.113***	***1.454***	***4.278***	0.598	**2.129**	***0.848***	**2.757**	***1.313***	***3.925***	***0.963***	***3.047***
Abundance CA axis 2	0.586	1.510	0.478	1.261	0.389	1.054	0.090	0.363	0.336	0.933	0.222	0.669
Abundance CA axis 3	0.060	0.121	0.236	0.579	-0.182	-0.511	-0.111	-0.326	0.057	0.113	-0.044	-0.151
P/A CA axis 1	***1.096***	***3.383***	***1.446***	***4.276***	0.629	**2.192**	0.411	1.637	***0.927***	**2.954**	0.557	**2.009**
P/A CA axis 2	0.187	0.540	0.205	0.585	0.479	1.292	-0.564	-1.399	0.126	0.381	0.343	0.941
P/A CA axis 3	-0.217	-0.607	-0.426	-1.162	0.041	0.076	0.230	0.578	-0.168	-0.477	0.071	0.155
Generalized distances to reference-area centroid												
3 abundance CA axes (95^{th} = 2.849; 99.9^{th} = 3.375)		***7.054***		***7.063***		***3.939***		***4.033***		***6.189***		***4.730***
3 P/A CA axes (95^{th} = 2.941; 99.9^{th} = 2.982)		***4.266***		***5.539***		***3.218***		1.918		***3.651***		2.739
11 two-tailed indices (95^{th} = 4.547; 99.9^{th} = 4.867)		***13.826***		***11.296***		***9.662***		***16.425***		***14.365***		***13.692***
associated F value (11 and 19 df)		341.58		228.00		166.79		482.04		368.71		334.97
p value (statistical significance; $\delta > 0$)		<0.001		<0.001		<0.001		<0.001		<0.001		<0.001
p value (ecological significance; $\delta > 2$ SDs)		<0.001		<0.001		<0.001		<0.001		<0.001		<0.001

CA scores ranked the test streams differently (Table 3). Two streams, CB1 and HP6, were not biologically unusual (i.e., were less than the 95th percentile of 2.94), whereas the other 4 streams were extreme, lying beyond the 99.9th percentile. Surprisingly, both multivariate assessments suggested that DE6 was unusual, although there was no indication of this when the 20 metrics were evaluated individually using the regional reference values (also see Figs 3A, 4A).

Generalized distances based on the 11 two-tailed indices revealed that all 6 test streams were extremely unusual, lying beyond the 99.9th percentile from the reference-stream centroid (Table 3). DE6 was the least deviant stream, followed by PC1, HP6, DE11 and HP3. This ranking differs from the gradient in stream pH, the 2 multivariate assessments using the individual sets of CA scores, and both evaluations based on all 20 metrics. The lack of agreement among the 3 tests underscores the need for an assessment strategy that unequivocally judges the biological condition of a test site.

Statistically judging impairment - The F values for the individual MANOVAs (or Hotelling's T^2) indicated that the generalized distances between each test stream and the centroid of the reference streams were all significantly >0 (Table 3, $p<0.001$). Thus, each of the test streams was significantly different from the centroid of the reference streams. Moreover, the non-central tests revealed that each test stream was significantly outside of the normal range defined by the reference streams ($p<0.001$). Thus, each test stream exhibited what might prove to be ecologically significant differences from the reference streams. This conclusion differs from our findings based on the other assessments discussed, although extreme values were observed for several test streams (Table 3).

The associated DFAs produced linear combinations of the 11 two-tailed metrics that maximally separated each test stream from the reference-stream centroid (Table 4). For PC1, all 11 t values associated with the DF axis were significantly >0 ($p<0.006$). However, the relatively small partial T^2 values indicated that many of the indices were redundant. This redundancy (or correlation) among the metrics produced a generalized distance for PC1 that was smaller than the corresponding distances for many of the other test streams (see Table 3). By contrast, CB1 was the most extreme test stream with a generalized distance of 16.4 where all but 2 indices contributed significantly to the associated DF axis (i.e., abundance CA axes 2 and 3, $t<2.0$, $p>0.05$, Table 4). But unlike PC1, a number of metrics contributed unique information to the discriminant axis (e.g., % chironomids with partial $T^2 = 6.01$).

Generalized distances for 3 test streams (HP3, HP6, and DE11) were quite similar (Table 3), although the DF axes indicated that high % dipterans distinguished HP6 from the other 2 streams (Table 4). The partial T^2 values also revealed that abundance CA axes 1 and 2, and P/A CA axes 1 and 3 generally contributed unique information to each DF axis. Abundance CA axis 3 and P/A CA axis 2 occasionally exhibited large t values, but the low partial T^2 values suggested that these 2 axes were redundant with other indices. These results emphasize that assessments which only focus on the total number of atypical or extreme values can potentially over-estimate the level of impairment

Table 4. Univariate t values and probabilities plus multivariate partial T^2 values for the 11 two-tailed indices associated with each test-stream discriminant analysis.

Index	DE11			PC1			DE6			CB1			HP3			HP6		
	t	p	Partial T^2	t	p	Partial T^2	t	p	Partial T^2	t	p	Partial T^2	t	p	Partial T^2	t	p	Partial T^2
% crustaceans	3.938	<0.001	1.049	3.938	<0.001	1.508	0.523	0.605	0.749	3.778	0.001	0.242	3.938	<0.001	0.558	3.777	0.001	0.073
% chironomids	0.317	0.754	2.541	8.045	<0.001	2.650	5.609	<0.001	5.029	7.059	<0.001	6.011	6.479	<0.001	3.398	0.989	0.331	4.056
% dipterans	1.985	0.057	1.907	3.895	0.001	1.697	1.711	0.098	3.741	4.094	<0.001	4.609	1.572	0.127	2.523	7.446	<0.001	3.619
% gastropods	3.070	0.005	2.074	3.070	0.005	1.340	3.070	0.005	2.681	3.070	0.005	1.539	3.070	0.005	1.822	2.922	0.007	2.517
% pelecypods	4.530	<0.001	1.393	4.530	<0.001	1.471	4.530	<0.001	0.912	3.281	0.003	0.908	0.988	0.331	1.511	4.029	<0.001	1.127
Abundance CA axis 1	22.530	<0.001	5.211	23.434	<0.001	3.015	11.662	<0.001	3.854	15.101	<0.001	5.408	21.497	<0.001	6.596	16.687	<0.001	8.349
Abundance CA axis 2	8.270	<0.001	4.737	6.907	<0.001	4.644	5.772	<0.001	3.242	1.989	0.056	3.163	5.111	<0.001	3.636	3.663	0.001	2.645
Abundance CA axis 3	0.661	0.514	1.678	3.173	0.004	1.330	2.797	0.009	2.551	1.785	0.085	0.974	0.617	0.542	1.577	0.825	0.416	1.884
P/A CA axis 1	18.530	<0.001	3.828	23.421	<0.001	2.066	12.008	<0.001	2.919	8.964	<0.001	4.591	16.181	<0.001	4.779	11.001	<0.001	5.929
P/A CA axis 2	2.958	0.006	0.092	3.206	0.003	0.282	7.077	<0.001	0.505	7.661	<0.001	1.049	2.088	0.046	0.035	5.154	<0.001	0.469
P/A CA axis 3	3.323	0.002	2.414	6.363	<0.001	1.758	0.417	0.680	1.738	3.166	0.004	3.256	2.610	0.014	2.342	0.848	0.404	2.199

assessments and we found that all 6 test streams were significantly different from the reference-stream centroid (Table 3). More importantly, we found that all 6 test streams were significantly outside of the normal range based on the reference streams. Traditional statistical tests are affected by power, sample-size, and effect-size issues (McBride et al. 1993, Underwood 1994), and thus, the interpretation of statistical significance is open to debate (Osenberg et al. 1994, Power et al. 1995). However, we found evidence that the biological condition of each test stream was sufficiently different from the reference streams to be potentially ecologically significant (sensu Kilgour et al. 1998). This finding leads the statistical-significance debate into a new arena where a linkage between a statistically significant non-central test and the ecological importance of the observed difference must be established (Kilgour et al. 1998).

Large numbers of Simuliidae and Plecoptera in the test streams contributed to significant differences relative to reference-stream communities. For example, only five Simuliidae were enumerated among the almost 3000 invertebrates in the reference-stream samples. By contrast, Simuliidae averaged 14-36% of the animals collected from the test streams. In 1981, Mackay and Kersey (1985) found that Simuliidae were widespread in acidic Dorset-area streams. Similar observations from a long-term study of acidic streams (Hall and Ide 1987), led Chmielewski and Hall (1993) to hypothesize that acid deposition had contributed to an increase in the relative abundances of Simuliidae in south-central Ontario streams.

Five of our six test streams were among the eight streams sampled by Mackay and Kersey (1985). They found that Trichoptera were the most abundant non-dipterans, whereas Ephemeroptera and Plecoptera were rare. Mackay and Kersey (1985) also reported that chironomids were abundant. We found that chironomids were common, representing from 9-42% of the invertebrates that we enumerated in the test streams. By contrast, Trichoptera averaged only 1-2% in the 2 Dickie Lake streams (DE6 and DE11), and ranged from 6-16% in the other test streams. We also found that Ephemeroptera were rare in 3 streams (i.e., 0-1% in DE11, HP3, and PC1), but ranged from 9-13% in the other test streams. Surprisingly, Plecoptera represented 30-73% of the animals that we collected from the individual sites in DE11, HP3, and PC1, although they were less common in DE6, HP6, and CB1, where Simuliidae were abundant (i.e., >30%). Simpson et al. (1985) found that Plecoptera can dominate the invertebrate community in acid-stressed streams; however, differences between our results and those of Mackay and Kersey (1985) may simply reflect differences in sampling methods (i.e., our kick and sweep vs leaf bags, artificial substrates, and Surber samples in their study). Alternatively, the invertebrate communities may have changed since their 1981 study. Whatever the reason, the benthic communities of the test streams are unusual with respect to the reference streams. Conceivably, this difference reflects the long-term acid-deposition history of the Dorset area (e.g., LaZerte and Dillon 1984, Molot et al. 1989).

Contrary to our expectations, we observed that the differences in biological condition were not consistent with the pH

of the test streams (Table 3). Generally the difference in biological condition depended on our choice of summary metrics and the assessment procedure. Given the low pH of the test streams (Table 1) and the relatively high abundances of Simuliidae, Plecoptera, and chironomids, we believe that the invertebrate communities of the test streams have not recovered from their historical impairment documented by Mackay and Kersey (1985). Although acid-deposition rates have declined in the Dorset area (Dillon and LaZerte 1992, Dillon et al. 1997), it is not clear how long it will take for the communities of the test streams to recover (e.g., Yan et al. 1996, Keller and Yan 1998).

Are regional reference values useful? Using the 95th percentile criterion, five of the six test streams were identified as lying outside of the normal range for at least one of the 20 indices. Surprisingly, only three of the metrics indicated potential impairment, and two of these indices were multivariate ordination axes. Multivariate metrics are widely recognized as being more powerful than simple metrics for identifying impacts (Warwick and Clarke 1991, Yan et al. 1996), but the apparent lack of impairment indicated by the majority of our indices is a concern.

The CA ordinations illustrate a common problem that can occur when biological condition is assessed by examining indices one at a time. On the 1st and 2nd axes of the abundance CA (Fig. 3A), 26 of the 30 reference samples produced an elliptical cluster indicating a large negative correlation between these 2 metrics. Test streams lying at either end of the major axis of this ellipse would be correctly identified as biologically unusual with either set of CA scores. However, test streams lying along the minor axis can fall outside of the ellipse, yet go undetected when examining each axis one at a time. As a result, a multivariate assessment that accounts for correlations (or covariation) among the metrics is warranted whenever several indices are evaluated (Norris 1995, Reynoldson et al. 1997). Many assessment methods ignore correlations among the metrics and thus, may have less power to detect impacts (Suter 1993, Norris 1995).

Our generalized distances (i.e., MANOVA) based on the 11 two-tailed indices indicated that all six test streams were statistically significantly outside of the normal range defined by the reference streams (Table 3). By contrast, the regional reference values suggested that the test streams ranged from normal (DE6) to extremely unusual for three of the 20 indices (DE11 and PC1). Individual metrics that contributed substantively to the DFA (i.e., t >2.0, Table 4), generally exhibited δs larger than 0.5 SDs (Table 3). For a 2-tailed metric, the $\bar{x} \pm 0.5$ SDs approximates the region enclosed by the 30th and 70th percentiles. This region is considerably smaller than our normal range defined by the 2.5 and 97.5 percentiles (Table 2), although it is similar to the quartiles (i.e., 25th and 75th percentiles) that were used by Gerritsen (1995) and Barbour et al. (1996). From our DFA results, large t values (i.e., t >15, Table 4) and δs >3 SDs (Table 3) were generally associated with test-stream metrics that fell outside of the regional reference values. Presumably, the conservative nature of our regional reference values reflects our interest in whether the test streams lie outside of the normal range, not simply that the observed difference is significantly different from 0.

The regional reference values also revealed problems with truncated distributions and a resultant lack of discriminating ability for several metrics (Table 2). In addition, the test streams exhibited an unexpected problem with several 1-tailed indices. That is, taxonomic richness, dipteran richness, % EPT, and % non-dipteran insects were often higher in the test streams than in the reference streams, and the % dominants were lower (Table 3). Thus, our initial assumption that 9 of our 20 indices were 1-tailed was not supported. Perhaps this problem arose because our indices were based on rapid bioassessment data with very coarse taxonomic resolution (Courtemanch 1996), or perhaps because the test streams are affected by multiple stressors (Schindler et al. 1996, Dillon et al. 1997). Alternatively, we based our expectations on responses to nutrient enrichment rather than acidification (e.g., Plafkin et al. 1989, Culp et al. 1997).

The observed increases in richness were probably due to the presence of Simuliidae, Tipulidae, and Ceratopogonidae in the test streams. Increases in the % EPT and % non-dipteran insects were undoubtedly due to large numbers of Plecoptera in the test streams. In hindsight, these patterns should have been anticipated (Mackay and Kersey 1985, Simpson et al. 1985, Chmielewski and Hall 1993) given the historical acid-deposition problem (LaZerte and Dillon 1984, Molot et al. 1989). As a result, we suggest that most indices should be considered to be 2-tailed (i.e., both low and high values may indicate unusual condition) unless the impacts of the stressor(s) in question are fully understood (also see Suter 1993).

Another explanation for the low power of most of our summary indices may lie in the fact that some of the reference streams were larger than the test streams. As a result, the reference-stream variation (i.e., SD_{ref}) may be larger than it should be, and our tentative regional reference values may be larger (or smaller) than they should be. Differences in the proportions of collectors and shredders are also expected when streams of different sizes are compared (Vannote et al. 1980, Townsend and Hildrew 1984, Grubaugh et al. 1996). Consequently, differences in stream order may have contributed to patterns that we have attributed to water-quality impairment. However, we assume that acidification, not stream order, is the major factor influencing our results because our observed differences are consistent with benthic communities that have been impaired by acidification (Mackay and Kersey 1985, Townsend and Hildrew 1984, Simpson et al. 1985). We also assume that the poor performance of most of our indices is due, in part, to our initial choice of indicators. Undoubtedly, metrics such as the % Simuliidae and % Plecoptera would indicate that the test streams are outside of the normal range.

Evaluating multiple indices simultaneously - Building upon the approach proposed by Kilgour et al. (1998), our generalized distances and associated MANOVAs provided a single assessment for each test stream (Table 3). The subsequent DFAs complement the generalized distances and identified the metrics that distinguished the test-stream communities (Table 4). Indices with the largest t values on a given DF axis were also flagged by the regional reference values. In addition, the partial T^2 values indicated if a given metric was

redundant with the other indices. Thus, the MANOVA and DFA provide a powerful set of tools for impact assessments: 1) a single judgment based on the results of all indices simultaneously; 2) a method that is designed to take correlations (or redundancy) among metrics into account automatically; 3) a method that provides an assessment of the relative importance of different indices for distinguishing a test site; and 4), the ability to assess the statistical significance of the biological condition of a test site.

Because MANOVA and DFA are traditional statistical procedures, there are concerns regarding the robustness of multivariate, hypothesis-testing approaches to non-normality and outliers (Gerritsen 1995, Norris 1995). In our study, both the abundance and P/A CAs indicated that 4 reference samples were unusual relative to the other 26 samples (Figs 3A, 4A). Despite this potential problem, the 1st and 2nd axes of both CAs were among the most important summary indices for distinguishing the test streams (Table 4). Alternatively, robust statistical estimators that minimize the effects of unusual observations could have been used (e.g., Campbell 1980, 1982). Given our observation that the frequency distributions for several indices from the reference streams were not normally distributed, further work in this area seems warranted.

Our MANOVA and DFA differ considerably from multimetric (e.g., Barbour et al. 1995) and percent model affinity (Novak and Bode 1992) approaches. The generalized distance down-weights redundant information, and thus, a large number of indices can be included in an assessment. Moroever, the DFA will identify those metrics that contribute to the distinct nature of a test site (e.g., Table 4). This feature avoids the labour intensive screening of candidate indices associated with the multimetric approach (e.g., Kerans and Karr 1994, Barbour et al. 1996), and avoids the calibration and construction of multimetric scoring thresholds (e.g., Minns et al. 1994, Barbour et al. 1995). As a result, the anticipated stressors and the stress-response relationships do not have to be known *a priori*. Assuming that a wide variety of indices are used, the MANOVA and DFA multivariate approach should be sensitive to the effects of a variety of stressors simultaneously.

Although MANOVA and DFA are multivariate methods, they are hypothesis-testing approaches that should not be confused with exploratory methods like ordination and clustering. Exploratory multivariate methods have been criticized for a number of reasons (Gerritsen 1995, Norris 1995, Fore et al. 1996), including a lack of significance testing, a multitude of competing methods, and unnecessary mathematical complexity. By contrast, our MANOVA involves a single significance test based on the simple concept of the biological distance between the center of the reference samples and the test site.

We also used exploratory ordination methods in this study. Our CA scores were used as multivariate summary indices reflecting benthic community composition, not unlike a % EPT or % dominants index. We evaluated the CA scores and a number of common indices; interestingly, the ordination scores proved to be among the most-sensitive indices for detecting unusual patterns in the test-stream communities. That is, the CA scores emphasized differences between

the test and reference-stream communities associated with the abundances and occurrences of Simuliidae, Plecoptera, and Tipulidae (Figs 3B, 4B) that were not captured by the other indices. Warwick and Clarke (1991) and Yan et al. (1996) also found that summary metrics based on exploratory multivariate methods were very sensitive to stress-induced changes in aquatic communities.

In this study, we used a rank-order version of the normal range to characterize the biological condition of a series of reference streams (e.g., David et al. 1998) and then evaluated 6 streams with a known history of water-quality impairment. Our results demonstrated that rapid bioassessment data can be used to characterize the biological condition of benthic communities, and that regional reference values can be used to evaluate the biological condition of potentially impacted streams. Because our regional-reference approach does not provide a single, numerical summary of the condition of a test site, we expanded on the ideas of Kilgour et al. (1998) and used MANOVA and DFA to statistically assess each of the 6 test streams. This multivariate assessment complements the results of the regional-reference approach, and emphasized the value of a single judgment based on a number of biological indices simultaneously. Using this statistical approach, the benthic communities of all 6 test streams were judged to be extremely unusual, and potentially biologically impaired. This finding suggests that the benthic invertebrate communities in these Dorset-area streams have not recovered from historical water-quality impairment.

CONCLUSIONS AND RECOMMENDATIONS

1) Rapid bioassessment data from regional reference sites can be used to assess the biological condition of test sites. The resultant regional reference values can be used to screen test sites, but we recommend that a MANOVA should be used to statistically judge the unusual nature of each test site. Both 1- and 2-tailed indices should be used in the analysis unless the stressors and their actions are fully understood. Because of their sensitivity to changes in the benthic community, ordination scores should be included as multivariate metrics in any assessment.

2) A significant MANOVA should be followed by a DFA to identify the indices that are most important for distinguishing a test site from the reference sites. Results of the DFA are readily interpreted with t and partial T^2 statistics. These follow-up comparisons provide an understanding of how a test site differs from the reference sites, and can be used to implicate a particular stressor or stressors.

3) Regional reference values based on the normal range of variation provide practical effect-size targets. Although a formal statistical test is easily implemented using the generalized distance and associated MANOVA, the biological or ecological significance of an impact that exceeds the normal range remains to be established. Future studies should explore the relationship between critical effect sizes based on the normal range and the ecological significance of environmental impacts.

ACKNOWLEDGEMENTS

Many colleagues encouraged us to evaluate the use of regional reference values in rapid bioassessments. We thank Peter Dillon, Neil Hutchinson, Don Jackson, Bruce Kilgour, Norm Yan, and Sheila David for their suggestions and support throughout this study. We also thank our colleagues and students who assisted with the field and laboratory work. Peter Dillon, Don Jackson, Bruce Kilgour, Jim Rusak, and Norm Yan provided constructive comments on the manuscript. Funding for this work was provided by the Ontario Ministry of Natural Resources and the Ontario Ministry of the Environment.

LITERATURE CITED

Aagaard, K., J.O. Solem, T. Nost, and O. Hanssen. 1997. The macrobenthos of the pristine stream, Skiftesaa, Hoylandet, Norway. Hydrobiologia 348:81-94.

Anonymous. 1983. Handbook of analytical methods for environmental samples. Volumes 1 and 2. 962 pp. Ontario Ministry of the Environment, Rexdale, Ontario.

Bailey, R.C., M.G. Kennedy, M.Z. Dervish, and R.M. Taylor. 1998. Biological assessment of freshwater ecosystems using a reference condition approach: comparing predicted and actual benthic invertebrate communities in Yukon streams. Freshwater Biology 39: 765-774.

Barbour, M.T., J. Gerritsen, G.E. Griffith, R. Frydenborg, E. McCarron, J.S. White, and M.L. Bastian. 1996. A framework for biological criteria for Florida streams using benthic macroinvertebrates. Journal of the North American Benthological Society 15:185-211.

Barbour, M.T., J.L. Plafkin, B.P. Bradley, C.G. Graves, and R.W. Wisseman. 1992. Evaluation of EPA's rapid bioassessment benthic metrics: metric redundancy and variability among reference stream sites. Environmental Toxicology and Chemistry 11:437-449.

Barbour, M.T., J.B. Stribling, and J.R. Karr. 1995. Multimetric approach for establishing biocriteria and measuring biological condition. Pages 63-77 *in* W.S. Davis and T.P. Simon (editors). Biological assessment and criteria, tools for water resource planning and decision making. Lewis Publishers, Boca Raton, Florida.

Barton, D.R. 1996. The use of the Percent Model Affinity to assess the effects of agriculture on benthic invertebrate communities in headwater streams of southern Ontario, Canada. Freshwater Biology 36:397-410.

Billheimer, D., T. Cardoso, E. Freeman, P. Guttorp, H.-W. Ko, and M. Silkey. 1997. Natural variability of benthic species composition in the Delaware Bay. Environmental and Ecological Statistics 4:95-115.

Campbell, N.A. 1980. Robust procedures in multivariate analysis. I: Robust covariance estimation. Applied Statistics 29:231-237.

Campbell, N.A. 1982. Robust procedures in multivariate analysis. II: Robust canonical variate analysis. Applied

Statistics 31:1-8.

Chessman, B.C. 1995. Rapid assessment of rivers using macroinvertebrates: a procedure based on habitat-specific sampling, family-level identification and a biotic index. Australian Journal of Ecology 20:122-129.

Chmielewski, C.M., and R.J. Hall. 1993. Changes in the emergence of blackflies (Diptera: Simuliidae) over 50 years from Algonquin Park streams: is acidification the cause? Canadian Journal of Fisheries and Aquatic Sciences 50:1517-1529.

Clair, T.A., J. Ehrman, and K. Higuchi. 1998. Changes to the runoff of Canadian ecozones under a doubled CO_2 atmosphere. Canadian Journal of Fisheries and Aquatic Sciences 55:2464-2477.

Clark, B. 1996. Limnological sample and data collection methods used by the Muskoka Lakes Aquatic Assessment Unit. Dorset Environmental Science Centre, Ontario Ministry of Environment and Energy, Dorset, Ontario. ISBN: 0-7778-5757-X.

Corkum, L.D. 1989. Patterns of benthic invertebrate assemblages in rivers of northwestern North America. Freshwater Biology 21:191-205.

Courtemanch, D.L. 1996. Commentary on the subsampling procedures used for rapid bioassessments. Journal of the North American Benthological Society 15:381-385.

Culp, J.M., K.J. Cash, and D.B. Halliwell. 1997. Volunteer-based monitoring program for the Salmon River basin: using benthic indicators to assess stream ecosystem health. National Hydrology Research Institute, Environment Canada, Saskatoon, Saskatchewan, 53 pp. + app.

David, S.M., K.M. Somers, R.A. Reid, R.J. Hall, and R.E. Girard. 1998. Sampling protocols for the rapid bioassessment of streams and lakes using benthic macroinvertebrates. 2nd edition, Dorset Environmental Science Centre, Ontario Ministry of the Environment, Dorset, Ontario. ISBN: 0-7778-6504-1.

Diamond, J.M., M.T. Barbour, and J.B. Stribling. 1996. Characterizing and comparing bioassessment methods and their results: a perspective. Journal of the North American Benthological Society 15:713-727.

Dillon, P.J., and B.D. LaZerte. 1992. Response of the Plastic Lake catchment, Ontario, to reduced sulphur deposition. Environmental Pollution 77:211-217.

Dillon, P.J., and L.A. Molot. 1997. Effect of landscape form on export of dissolved organic carbon, iron, and phosphorus from forested stream catchments. Water Resources Research 33:2591-2600.

Dillon, P.J., L.W. Molot, and M. Futter. 1997. The effect of El Nino-related drought on the recovery of acidified lakes. Environmental Monitoring and Assessment 46:105-111.

Dillon, P.J., L.W. Molot, and W.A. Scheider. 1991. Phosphorus and nitrogen export from forested stream catchments in central Ontario. Journal of Environmental Quality 20:857-864.

Fisher, R.A., A.S. Corbet, and C.B. Williams. 1943. The relation between the number of species and the number of individuals in a random sample of an animal

population. Journal of Animal Ecology 12: 42-58.

Fore, L.S., J.R. Karr, and L.L. Conquest. 1996. Assessing invertebrate responses to human activities: evaluating alternative approaches. Journal of the North American Benthological Society 15:212-231.

Gerritsen, J. 1995. Additive biological indices for resource management. Journal of the North American Benthological Society 14:451-457.

Girard, R., and R.A. Reid. 1990. Dorset Research Centre study lakes: sampling methodology (1986-1989) and lake morphometry. Dorset Research Centre, Ontario Ministry of the Environment, Dorset, Ontario. 33 pp. + app.

Glasby, T.M. 1997. Analyzing data from post-impact studies using asymmetrical analyses of variance: a case study of epibiota on marinas. Australian Journal of Ecology 22:448-459.

Growns, J.E., B.C. Chessman, J.E. Jackson, and D.G. Ross. 1997. Rapid assessment of Australian rivers using macroinvertebrates: cost and efficiency of 6 methods of sample processing. Journal of the North American Benthological Society 16:682-693.

Grubaugh, J.W., J.B. Wallace, and E.S. Houston. 1996. Longitudinal changes of macroinvertebrate communities along an Appalachian stream continuum. Canadian Journal of Fisheries and Aquatic Sciences 53:896-909.

Hall, R.J., and J.P. Ide. 1987. Evidence of acidification effects on stream insect communities in central Ontario between 1937 and 1985. Canadian Journal of Fisheries and Aquatic Sciences 44:1652-1657.

Hannaford, M.J., and V.H. Resh. 1995. Variability in macroinvertebrate rapid-bioassessment surveys and habitat assessments in a northern California stream. Journal of the North American Benthological Society 14:430-439.

Hartig, J.H., M.A. Zarull, T.B. Reynoldson, G. Mikol, V.A. Harris, R.G. Randall, and V.W. Cairns. 1997. Quantifying targets for rehabilitating degraded areas of the Great Lakes. Environmental Management 21:713-723.

Hughes, R.M. 1995. Defining acceptable biological status by comparing with reference conditions. Pages 31-48 *in* W.S. Davis and T.P. Simon (editors). Biological assessment and criteria, tools for water resource planning and decision making. Lewis Publishers, Boca Raton, Florida.

Hughes, R.M., D.P. Larsen, and J.M. Omernik. 1986. Regional reference sites: a method for assessing stream potentials. Environmental Management 10:629-635.

Jeffries, D.S., and W. Snyder. 1983. Geology and geochemistry of the Muskoka-Haliburton study area. Dorset Research Centre, Ontario Ministry of the Environment, Dorset, Ontario. 101 pp.

Keller, W., and N.D. Yan. 1998. Biological recovery from lake acidification: zooplankton communities as a model of

patterns and processes. Restoration Ecology 6:364-375.

Kerans, B.I., and J.R. Karr. 1994. A benthic index of biotic integrity (B-IBI) for rivers of the Tennessee Valley. Ecological Applications 4:768-785.

Kersting, K. 1991. Microecosystem state and its response to the introduction of a pesticide. Internationale Vereinigung fur Theoretische und Angewandte Limnologie Verhandlungen 24: 2309-2312.

Kilgour, B.W., K.M. Somers, and D.E. Matthews. 1998. Using the normal range as a criterion for ecological significance in environmental monitoring and assessment. Ecoscience 5:542-550.

LaZerte, B.D., and P.J. Dillon. 1984. Relative importance of anthropogenic versus natural sources of acidity in lakes and streams of central Ontario. Canadian Journal of Fisheries and Aquatic Sciences 41:1664-1677.

Leffler, J.W. 1978. Ecosystem responses to stress in aquatic microcosms. Pages 102-119 *in* J.H. Thorp and J.W. Gibbons (editors). Energy and environmental stress in aquatic systems. US DOE Symposium Series No. 48, Augusta, Georgia.

Lenat, D.R., and M.T. Barbour. 1994. Using benthic macroinvertebrate community structure for rapid, cost-effective, water quality monitoring: rapid bioassessment. Pages 187-215 *in* S.L. Loeb and A. Spacie (editors). Biological monitoring of aquatic systems. Lewis Publishers, Boca Raton, Florida.

Mackay, R.J., and K.E. Kersey. 1985. A preliminary study of aquatic insect communities and leaf decomposition in acid streams near Dorset, Ontario. Hydrobiologia 122:3-11.

Manly, B.F.J. 1986. Multivariate statistical methods. A primer. Chapman and Hall, New York.

McBride, G.B., J.C. Loftis, and N.C. Adkins. 1993. What do significance tests really tell us about the environment? Environmental Management 17:423-432.

Minns, C.K., V.W. Cairns, R.G. Randall, and J.E. Moore. 1994. An Index of Biotic Integrity (IBI) for fish assemblages in the littoral zone of Great Lakes' Areas of Concern. Canadian Journal of Fisheries and Aquatic Sciences 51:1804-1822.

Molot, L.A., P.J. Dillon, and B.D. LaZerte. 1989. Factors affecting alkalinity concentrations of streamwater during snowmelt in central Ontario. Canadian Journal of Fisheries and Aquatic Sciences 46:1658-1666.

Norris, R.H. 1995. Biological monitoring: the dilemma of data analysis. Journal of the North American Benthological Society 14:440-450.

Novak, M.A., and R.W. Bode. 1992. Percent model affinity: a new measure of macroinvertebrate community composition. Journal of the North American Benthological Society 11:80-85.

Osenberg, C.W., R.J. Schmitt, S.J. Holbrook, K.E. Abu-Saba, and A.R. Flegal. 1994. Detection of environmental impacts: natural variability, effect size, and power analysis. Ecological Applications 4:16-30.

Plafkin, J.L., M.T. Barbour, K.D. Porter, S.K. Gross, and R.M. Hughes. 1989. Rapid bioassessment protocols for use in streams and

rivers: Benthic macroinvertebrates and fish. EPA/440/4-89/001. Office of Water Regulations and Standards, US EPA, Washington, DC.

Power, M., D.G. Dixon, and G. Power. 1995. Detection and decision-making in environmental effects monitoring. Environmental Management 19:629-639.

Rencher, A.C. 1992. Interpretation of canonical discriminant functions, canonical variates, and principal components. The American Statistician 46:217-225.

Rencher, A.C., and D.T. Scott. 1990. Assessing the contribution of individual variables following rejection of a multivariate hypothesis. Communications in Statistics Part B: Simulation and Computation 19:535-553.

Resh, V.H., R.H. Norris, and M.T. Barbour. 1995. Design and implementation of rapid assessment approaches for water resource monitoring using benthic macroinvertebrates. Australian Journal of Ecology 20:108-121.

Reynoldson, T.B., R.H. Norris, V.H. Resh, K.E. Day, and D.M. Rosenberg. 1997. The reference condition: a comparison of multimetric and multivariate approaches to assess water-quality impairment using benthic macroinvertebrates. Journal of the North American Benthological Society 16:833-852.

Rohlf, F.J. 1993. NTSYS-pc numerical taxonomy and multivariate analysis system. Version 1.8. Exeter Publishing Ltd., Setauket, New York.

Roux, D.J., P.L. Kempster, C.J. Kleynhans, H.R. van Vliet, and H.H. du Preez. 1999. Integrating stressor and response monitoring into a resource-based water-quality assessment framework. Environmental Management 23:15-30.

Schindler, D.W. 1998. A dim future for boreal waters and landscapes. BioScience 48:157-164.

Schindler, D.W., S.E. Bayley, B.R. Parker, K.G. Beaty, D.R. Cruikshank, E.J. Fee, E.U. Schnidler, and M.P. Stainton. 1996. The effects of climatic warming on the properties of boreal lakes and streams at the Experimental Lakes Area, northwestern Ontario. Limnology and Oceanography 41:1004-1017.

Simpson, K.W., R.W. Bode, and J.R. Colquhoun. 1985. The macroinvertebrate fauna of an acid-stressed headwater stream system in the Adirondack Mountains, New York. Freshwater Biology 15:671-681.

Sneath, P.H.A., and R.R. Sokal. 1973. Numerical taxonomy. The principles and practice of numerical classification. W.H. Freeman and Company, San Francisco.

Snyder, B.D., J.B. Stribling, M.T. Barbour, and C.L. Missimer. 1999. Integrating assessments of fish and macroinvertebrate assemblages and physical habitat condition in Pennsylvania. Pages 639-652 *in* T.P. Simon (editor). Assessing the sustainability and biological integrity of water resources using fish communities. CRC Press, Boca Raton, Florida.

Somers, K.M., R.A. Reid, and S.M. David.

1998. Rapid biological assessments: how many animals are enough? Journal of the North American Benthological Society 17:348-358.

Suter, G.W. 1993. A critique of ecosystem health concepts and indexes. Environmental Toxicology and Chemistry 12:1533-1539.

ter Braak, C.J.F., and I.C. Prentice. 1988. A theory of gradient analysis. Advances in Ecological Research 18:271-317.

Townsend, C.R., and A.G. Hildrew. 1984. Longitudinal pattern in detritivore communities of acid streams: a consideration of alternative hypotheses. Verhandlungen Internationale Vereinigung fur Theoretische und Angewandte Limnologie 22:1953-1958.

Underwood, A.J. 1992. Beyond BACI: the detection of environmental impacts on populations in the real, but variable, world. Journal of Experimental Marine Biology and Ecology 161:145-178.

Underwood, A.J. 1994. On beyond BACI: sampling designs that might reliably detect environmental disturbances. Ecological Applications 4:3-15.

Vannote, R.L., G.W. Minshall, K.W. Cummins, J.R. Sedell, and C.E. Cushing. 1980. The river continuum concept. Canadian Journal of Fisheries and Aquatic Sciences 37:130-137.

Warwick, R.M., and K.R. Clarke. 1991. A comparison of some methods for analyzing changes in benthic community structure. Journal of the Marine Biological Association of the United Kingdom 71:225-244.

Yan, N.D., W. Keller, K.M. Somers, T.W. Pawson, and R.E. Girard. 1996. Recovery of crustacean zooplankton communities from acid and metal contamination: comparing manipulated and reference lakes. Canadian Journal of Fisheries and Aquatic Sciences 53:1301-1327.

Yoder, C.O., and E.T. Rankin. 1995. Biological criteria program development and implementation in Ohio. Pages 109-144 *in* W.S. Davis and T.P. Simon (editors). Biological assessment and criteria, tools for water resource planning and decision making. Lewis Publishers, Boca Raton, Florida